Subtle Cyber-Discrimination? Not *if* but *how* internet housing providers respond differently to Neil, Tyrone and Jorge

September 16, 2010

Maria Krysan

Department of Sociology and Institute of Government and Public Affairs
University of Illinois at Chicago

Samantha Friedman

Department of Sociology
University at Albany, SUNY

Gregory D. Squires

Department of Sociology
George Washington University

ABSTRACT

The study of racial and ethnic discrimination in today's housing market has become increasingly difficult because the nature of the housing industry has changed and the way households search for housing involves the use of the internet. Little research has captured these changes. One recent study that conducted correspondence tests of housing providers who advertised rental units on Craigslist in Boston and Dallas finds that large majorities of testers, regardless of the race or ethnicity implied by their names, receive responses from housing providers (Friedman et al. 2010). The high response rate to the testers raises the issue of whether qualitative differences exist in the nature of the responses to the testers on the basis of their race or ethnicity. This study will be the first qualitative analysis to systematically examine actual e-mail messages exchanged between testers and housing providers and document potential racial and ethnic differences in treatment.

INTRODUCTION AND BACKGROUND

Forty years after the Federal Fair Housing Act was passed, metropolitan areas continue to be highly segregated. Although there have been some declines in recent years, black-white segregation remains particularly high, and Hispanic/white and Asian/white segregation has actually increased somewhat (Iceland et al. 2002). It is generally recognized that discrimination in the housing market, despite being illegal, continues to be a major cause of these persistent patterns of segregation. The Housing Discrimination Studies (HDS), conducted by HUD in 1989 and 2000, have been an important source of data on trends in housing discrimination. These studies reveal that levels of discrimination have declined somewhat, but it is still the case that minority homeseekers are treated less favorably than white homeseekers (Turner et al. 2002).

These studies, however, have been criticized by scholars, policymakers and fair housing advocates (including some of the authors of these studies) for likely under-estimating the prevalence of discrimination because of the methodology used. One of the main limitations is that discrimination is estimated based only on a 'complete' audit—one in which the white, Latino, and African American tester were all able to meet with the housing provider. This means that if discrimination occurred in a given audit at the level of 'returning a phone call' (or not) or 'setting up an appointment" (or not) then that audit was not included in the analysis. Massey and colleagues (2001, 2004), in a telephone-based housing audit study demonstrated that discrimination (what they called 'linguistic profiling') can and does occur before a face-to-face meeting even occurs between housing providers and potential tenants.

Aside from the Massey studies, however, no research has been disseminated among scholars or policymakers that examine the discrimination on the part of real estate agents that appears to be occurring before the initial visit to the agent or home. In addition, there has been little

research on the extent to which these processes occur—or not—in the rapidly growing electronically advertised housing market in the U.S. (Carpusor and Loges 2006; Ewens et al. 2009). Of particular interest is the internet study by Friedman et al. (2010) because it is the only US-based study that considers treatment of testers by the same housing provider and focuses on the treatment of Hispanics in addition to blacks. Between January and May of 2009, Friedman et al. (2010) conducted an audit study of the electronic rental market in the Boston and Dallas metropolitan areas. A random sample of landlords from rental units advertised via Craigslist received essentially identical inquiries from men with white-, black- and Latino-sounding names. A total of 1,467 housing providers were tested resulting in 4,401 e-mail contacts. The goal was to determine if the responses received by these three auditors were influenced by the race/ethnicity of the auditor. Friedman et al. (2010) report the first set of findings from this audit study and two findings are of particular interest to the present analysis.

First, they report quite high levels of responses from the housing providers: more than 3 out of 4 auditors—regardless of race/ethnicity—were given the courtesy of a response to an auditor's email inquiry. The authors observe that this is a strikingly high percentage. Second, there is nevertheless evidence of disparate treatment—the white tester, on a number of outcomes, is given better treatment than the black tester and Latino tester. The outcomes included whether or not: the tester received a response; they received more than one response; the unit was available; the auditor was invited to inspect the unit; and the auditor was told to contact the provider.

These 'objective' indicators of how the auditors were treated are a useful barometer of disparate treatment. And they address the fundamental question of "if" Tyrone, Jorge, and Neil are treated differently. However, as scholars of discrimination have noted, discrimination is a 'moving target' (Massey 2005) and has become more subtle than in past eras, when

homeseekers might simply have been told “Negroes need not apply”. The more contemporary forms of discrimination take the form of ‘cold shoulders,’ less friendly/polite behavior, and more complicated barriers (e.g., more pre-screening requirements) to minority housing seekers than to white housing seekers.

As Friedman et al. (2010) point out in the conclusion of their paper, their analysis is limited to quantitative outcomes of access to housing. They find that most testers receive a response and are told the unit is available. But the quantitative analysis fails to determine if the quality of that response is the same across testers. What appears to be a “similar” response in a quantitative analysis might, upon a closer qualitative investigation, prove to be disparate treatment. There are many ways that the response can differ. For example, one tester may be more strongly encouraged by the landlord to view the home, with a lot of friendliness and enthusiasm; while another receives an abrupt response that does not encourage further communication. Or one tester might be given different options for viewing the home, and provided with many additional details about the unit—in an attempt to ‘sell’ the unit; and another tester is simply told it is available. In these cases all auditors would be coded as “invited to inspect” but the housing provider is clearly sending different messages. These kinds of differences can be identified through a detailed analysis of the text of the email responses from the housing providers but escape notice in less qualitatively oriented approaches.

Due to the nature of this housing audit research design, we are in a position to pursue these ideas through additional analysis of the Friedman et al. (2010) audit study. Most housing audit research designs rely on post-experience questionnaires completed by housing auditors to ascertain the nature of the interaction they had with the auditor. In the case of our electronic housing market audit study, which relies on email communication, we have the exact responses from the housing providers. Thus, we have an opportunity to explore in a more qualitative

fashion the possibility of the more subtle discrimination that is the hallmark of contemporary discrimination.

We do just that in this paper. Specifically, we analyze the text of the responses the auditors received to their inquiries into available housing units to ascertain if there are differences in the tone, quality, or content of the response, and if these differences constitute disparate treatment and perhaps unlawful behavior. For example, in some responses the housing provider expresses great enthusiasm about meeting the homeseeker, showing their units, and offering to do so at any time. In others the homeseeker receives a more abrupt response with only very limited times that the unit can be inspected. So both auditors receive a response and the housing provider is willing to show the unit; but the housing provider is clearly more interested in renting to one auditor than the other. This study, then, allows us to investigate not just "if" an auditor got a response—for as Friedman et al. report (2010), most auditors did. Here we are able to gauge "how" the auditors were responded to—and whether there were differences in how they were treated based on whether the name of the inquirer sounded African American, White, or Latino.

DATA AND METHODS

This study uses the original Friedman et al. (2010) internet housing provider audit study of a random sample of Boston and Dallas rental housing providers who posted advertisements via Craigslist. As mentioned above, the 1,467 rental housing ads received a total of 4,401 e-mail contacts from the three auditors. For the purposes of this study, we analyzed the content of every email that was received by the auditors. The responses from housing providers were analyzed using two different coding approaches. For the first, we were interested in capturing as much of the detail and nuance of the email messages as possible. For Round 1 of our

coding, we relied on open-ended coding techniques that maximized retention of the information from the original text of the audits. A complex coding scheme was developed that permitted coding of all themes that emerged in the data. There was no limit to the number of codes that could be used, and the coding categories were not mutually exclusive.

Two graduate research assistants, working with the first author, developed the coding scheme used in this analysis. The process for creating the coding scheme was both deductive and inductive; and therefore iterative. Based on existing understanding of how racial discrimination occurs in the housing market, an initial set of coding categories (themes) was created. This coding scheme was used to test-code subsets of the audit data (with samples drawn from both metropolitan areas) and the coding scheme was modified several times to include additional themes that emerged and to provide more descriptive information about the themes so as to increase the clarity and specificity of the coding scheme. This, in turn, resulted in a coding scheme that could be used with a satisfactory level of inter-coder reliability.

During production coding, each email audit was carefully read and coded into the different themes, which were designed to gauge “how” the housing providers responded. The coding scheme included codes for the level of formal and grammatically correct language; the presence and formality of the salutation/greeting and the closing; the use of adjectives and modifiers to convey feelings of enthusiasm and friendliness; presence of accommodations toward the auditor, as well as existence of restrictions for the auditor; whether the auditor was invited to contact the provider (or vice versa); the presence and type of questions posed to the auditor about their needs, preferences, or experiences; the offering of information that increased the attractiveness of the housing unit (e.g., efforts to ‘sell’ the place to the auditor) and that decreased its attractiveness; the presence of questions posed by the agent to ascertain the

social class of the auditor; additional information provided to the auditor; and inquiries by the agent about additional information from the auditor.

Four research assistants were responsible for completing the production coding for Round 1. The RAs were trained in the coding scheme (and involved in making modifications to it) and a series of reliability comparisons were completed prior to production coding. There is no standard measure of inter-coder reliability, and our use of a complex coding scheme that was also not mutually exclusive made a measure of inter-coder reliability difficult to obtain. We used two approaches. First, we calculated a Kappa statistic for each theme (Siegel and Castellan 1988). The Kappa statistic calculates the level of agreement between two independent coders that takes into account the possibility that their agreement could be due to chance, rather than to a reliable measurement instrument. Due to the complexity of our coding scheme, and the features of the Kappa statistic, where there are small cell sizes (which occur quite frequently in our complex coding scheme), Kappa itself becomes an unreliable indicator of agreement rates. As such, reliability was also gauged using a simple percentage agreement. Specifically, for each of the themes included in the coding scheme, we calculated the percentage agreement (the percentage of times all four coders agreed on the coding categories exactly). We proceeded to production coding once the Kappa statistics across coders, and across each theme, exceeded .75 and/or the simple 'percent agreement' exceeded 80 percent. Once reliability was achieved across the four research assistants, each coder was assigned a group of audits to code, including audits from early and late in the field period and also including Boston and Dallas audits to ensure that there were no systematic differences in the sample of cases any particular RA was responsible for coding.

Our Round 2 coding approach was quite different from Round 1's very detailed analysis of the content of the responses received by each auditor. Round 2 involved ascertaining a global and

relative assessment of the response (or responses) each auditor received. Specifically, for each housing provider who was sent three email inquiries (one by each tester) and who responded to at least two of the testers, we assessed whether the responses, taken as a whole (e.g., if the tester received multiple responses, the entire set of responses were taken as a set and judged relative to the entire set of responses the other auditor(s) received), were more or less favorable to any particular auditor. To do this, we first 'redacted' all of the auditor/housing provider email exchanges to remove any mention of the name of the auditor (so as not to bias the coders in their assessments). A simple "1, 2, 3" designation[1] was used to identify the auditor who received the 'best' (1) treatment, the 'second best' treatment (2) and the 'worst' treatment (3).[2] Because of the greater degree of subjectivity in the Round 2 coding approach, we double coded each provider response (two coders independently coded each response). However, we first established a degree of agreement across all five coders (four graduate research assistants and the first author). Specifically, all five coders coded the same set of cases. Once the five coders had achieved a minimum of 80% agreement on the rank order of the auditors' responses across the five coders and two coding teams, we moved to production coding. The coders were divided into two teams (one two-member team; one three-member team). Each auditor/housing provider exchange was coded, independently, by two coders. The codes were compared, disagreements identified, and coders discussed and resolved all disagreements.

[1] The coding scheme appears deceptively simple. However, a number of coding decisions were agreed upon prior to implementing the production coding. This included what to do with 'identical' or 'near-identical' responses, and whether to give 'priority' to certain different kinds of treatment over other kinds of treatment. Our decision rules are outlined in the attached Appendix.

[2] If just two of the three auditors were given a response, then we ranked the responses 1 (best treatment) and 2 (worst treatment).

ANALYSIS APPROACH

For Round 1 Coding, each of the coding categories or themes can be understood as a binary variable (the theme was either present or absent in the housing provider's response). The coding scheme therefore translates into a number of possible indicators of disparate treatment. First, we will document whether or not the qualitative aspects of the responses to the housing inquiries differed for those with white, black, or Latino-sounding names and the frequency with which such differences occur across the audits. If whites, blacks, and Latinos differ in the extent to which a particular theme is reflected in the text of the audit, we will have evidence of disparate treatment.

Second, we will conduct multivariate logistic regression analysis to clarify and make more precise how race/ethnicity influences the experience of housing search in the electronic housing market. Although the audit methodology is experimental in nature, there are many factors that are not controlled for from audit to audit (e.g. ,whether the housing provider is from a real estate agent or an individual renting his or her own property directly; characteristics of the advertised unit; order of e-mail contacts; day of week/month of audit, etc.). Thus, we will conduct multivariate regression analyses that control for these variables in order to more precisely estimate the effect of race/ethnicity on the qualitative aspects of the online housing search process. Because the three observations within each audit will have the same characteristics, a multivariate logistic regression procedure that adjusts for within-audit correlation thereby provides robust standard error estimates will be used. Such an adjustment is necessary so as not to overstate the statistical significance of the results. Finally, we will compare the findings for Dallas and Boston in the descriptive and multivariate analyses outlined above.

We will also conduct analyses using the Round 2 coding procedure (the global indicator of disparate treatment). Frequencies will be reported for this variable, revealing the number of audits coded as identical treatment, white favored, African American favored, Latino favored, or mixed. In addition, multinomial logistic regression analyses will be conducted on this multi-category, global indicator of disparate treatment to identify factors that are associated with such disparities and which influences variations in access to housing for homeseekers with white-, black-, and Latino-sounding names in the electronic housing market. Finally, the descriptive and multivariate analyses will be compared between Boston and Dallas.

RESULTS AND CONCLUSIONS

At this point, the data have been coded and we are in the final stages of data cleaning. Once the data are ready, we will conduct our statistical analyses.

REFERENCES

Carpusor, Adrian G. and William E. Loges. 2006. “Rental Discrimination and Ethnicity in Names,” *Journal of Applied Social Psychology* 36 (4): 934-952.

Ewens, Michael, Bryan Tomlin, and Choon Wang. 2009. “Statistical Discrimination in the U.S. Apartment Rental Market: A Large Sample Field Study.” Unpublished paper.

Fischer, Mary J. and Douglas S. Massey. 2004. “The Ecology of Racial Discrimination.” *City and Community* 3(3):221-241.

Friedman, Samantha, Gregory D. Squires, and Chris Galvan. 2010. “Cyberdiscrimination in Boston and Dallas: Is Neil a More Desirable Tenant than Tyrone or Jorge?” Unpublished manuscript.

Iceland, John, Daniel H. Weinberg, and Erika Steinmetz. 2002. U.S. Census Bureau, Series CENSR-3, *Racial and Ethnic Residential Segregation in the United States: 1980-2000*. Washington, DC: U.S. Government Printing Office.

Massey, Douglas S. 2005. “Racial Discrimination in Housing: A Moving Target.” *Social Problems* 52(2): 148-151.

Massey, Douglas S. and Garvey Lundy. 2001. “Use of Black English and Racial Discrimination in Urban Housing Markets: New Methods and Findings.” *Urban Affairs Review* 36(4):452-469.

Siegel, Sidney, and N.J. Castellan, Jr. 1988. Nonparametric Statistics for the Behavioral Sciences. Second Edition. McGraw-Hill.

Turner, Margery Austin, Stephen L. Ross, George C. Galster, and John Yinger. 2002. *Discrimination in Metropolitan Housing Markets: National Results from Phase I HDS 2000.* Washington, DC: The Urban Institute.

APPENDIX A

Round 2 Coding Guidelines

1.) Our general principle for establishing rank order was to avoid, as much as possible, 'ties'. However, it was also the case that often housing providers 'cut and pasted' responses across auditors (i.e., they received identical—or virtually identical—responses). Our coding system included a designation for establishing that one or more response was 'identical' to the other. This included both 'word for word' identical responses, as well as responses that may have varied slightly in language (a few words here or there) but were for all practical purposes an identical response. However, if the response had just a few words difference, but those different words conveyed a quite different tone, then they were not permitted to be coded as "identical". Example of cases that were not 'word for word' identical, but were treated as 'identical' might be, for example, "When can you stop by to see it?" versus "When are you able to stop by and see it?"

2.) Sometimes one response is 'friendlier' in tone (or apparent tone) while the other response is more 'useful'. That is, one might say, "I'd love to show you the place" and another might say, "I can show you the place this Saturday at 3pm or anytime that works for you". While the former is more 'friendly,' the latter provides more information and accommodations to the auditor. The housing provider is essentially 'moving the renting process along' more aggressively. We made the decision that 'friendliness' could not trump the concrete gestures/information/etc. that facilitates the rental process. In the example here, the latter, while lacking the 'friendliness' gesture, would get coded as the 'better' response. The same principle applies for the statement "It is available". Specifically, if the only difference between two audits is that one says explicitly that "It is available" and the other does not use these specific words, but the responses makes it clear that the unit is, in fact, available (e.g., offers to show it to the auditor the next day), then the two audits would be coded as 'identical'. In other words, the statement "it is available"

cannot be the *only* reason that a particular audit is coded as 'better' than another. Finally, because the content of the email inquiries from the auditors varied systematically on whether or not the auditor asked about utilities (Tyrone/Tremayne and Jorge/Pedro both sometimes asked about it; Neil/Matthew never did) we ignored all mentions of utilities in the email response.